Using systematic observation to measure physical activity levels of primary school children during break-time: Issues, challenges and insight from an observational case study.

Michael Graham[1], Matthew Wright[1], Dr Alison Innerd[1], Dr Liane Azevedo[2], Prof. Alan Batterham[2]

@mikegraham84

1. Department of Sport and Exercise Sciences, School of Social Sciences, Humanities and Law, Teesside University, Middlesbrough, UK; 2. Health and Social Care Institute, Teesside University, Middlesbrough UK..

Brief overview

The purpose of the primary study was to assess current practices within a primary school playground within a culturally diverse community. In order to get an accurate measure of whole school physical activity levels, it was decided that systematic observation (SO) would give the most accurate representation of whole school physical activity levels.

SO often remains overlooked as a viable method for conducting research on PA behaviour.[1] Specific SO processes and protocols may be unfamiliar to researchers and deemed too labour intensive. This poster will focus on providing insight into the extensive process of training observers, recording and ensuring reliability and agreement for the use of SOPLAY, and the benefits of using video recordings.

Methods

Observation method

The System for Observing Play and Leisure Activity in Youth (SOPLAY) was selected for this study. SOPLAY is based on momentary time sampling techniques in which systematic and periodic scans of individuals and contextual factors within pre-determined target areas are made.[2] Any pupil within a specified target area during a scan was identified as actively participating and scored accordingly. Video recordings of pre-determined scan areas were made at regular intervals throughout an observation window of time and scored retrospectively (For a detailed description of systematic observation and SOPLAY see McKenzie 2016).[2]

SOPLAY Training

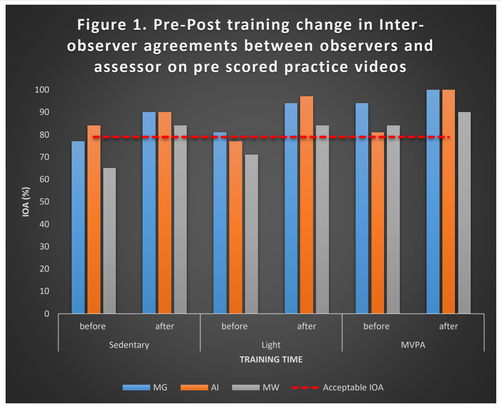
Observers spent a number of weeks using practice videos (available from Active Living Research https://hwb.cnr.ncsu.edu/resources/) and recording reliability data. Inter-observer agreement (IOA) and intra-class correlation coefficient (ICC 2,1)[3] were calculated to establish agreements between observers and the gold standard score. Inter-rater reliability was calculated using Cohens Kappa (k) with qualitative inference based on the following; >0.90, almost perfect; 0.80-0.90, strong; 0.60-0.79, moderate; 0.40-0.59, weak; 0.21-0.39, minimal; <0.20, none.[4] Practice videos were scored independently then discussed as a team. Only once observers met the criteria for acceptable reliability (IOA>80% and ICC>0.8) for the training and gold standard assessment video were they then able to move to the next stage of training (see figure 5. for the stages of training and calibration procedures for observers)**.**

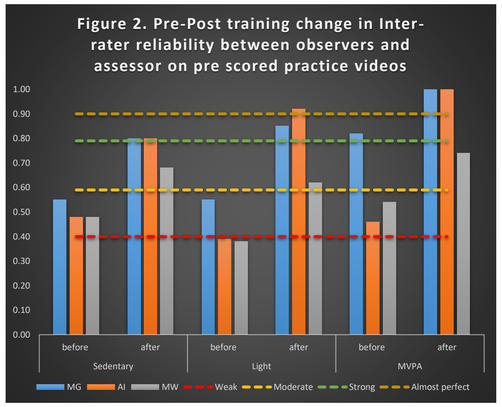
Pre recorded video

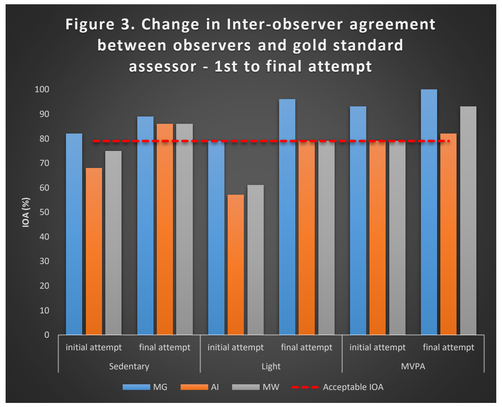
Six observers were trained in the mapping protocols for SOPLAY and were familiarised to the school playground and the selected target areas. The research team lead engaged with the school staff (teachers and playground supervisors) to ensure the target areas did not cross any boundaries or restrictions enforced by the school. Recordings were made at regular intervals throughout an observation window of time with observers working in pairs. Each pair started recording at the same time but at a different scan area and worked sequentially (Pair 1- area 1, Pair 2 – area 5, Pair 3 – area 10) and ended once break time was over.

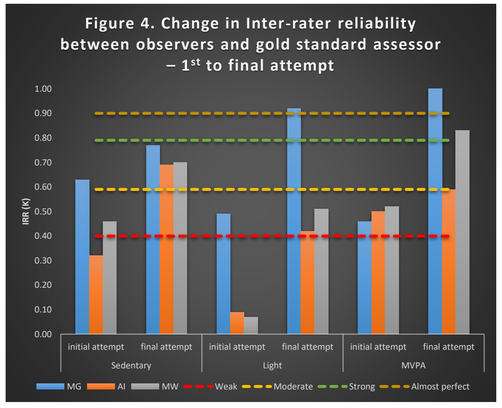
Reliability of observers –

The ICC for the practice videos were classed as 'very high' to 'extremely high' (Table 1.) based on the following thresholds: >0.99, extremely high; 0.99–0.90, very high; 0.75–0.90, high; 0.50–0.75, moderate; 0.20–0.50, low; <0.20, very low.[5] The ICC for the assessment videos were classed as 'very high' to 'extremely high' (table 2.) (0.93-1.00). IRR and IOA are presented in figures 1-4.









	Sedentary	Light	MVPA
Observer ID	ICC; (95%CI)		
MG	0.98; (0.96-0.99)	1.00; (0.99-1.00)	1.00; (1.00-1.00)
AI	0.91; (0.82-0.95)	0.93; (0.87-0.97)	0.94; (0.87-0.97)
MW	0.94; (0.87-0.97)	0.97; (0.94-0.99)	0.89; (0.76-0.94)

Table 1. Intra-Class Correlation Coefficient: Observer and assessor agreement for practice videos

	Sedentary	Light	MVPA
Observer ID	ICC; (95%CI)		
MG	0.97; (0.93-0.98)	0.99; (0.97-1.00)	1.00; (1.00)
AI	0.95; (0.87-0.97)	0.93; (0.84-0.96)	0.95; (0.87-0.97)
MW	0.96; (0.91-0.98)	0.96; (0.92-0.98)	0.98; (0.96-0.99)

Table 2. Intra-Class Correlation Coefficient: Observer and assessor agreement for assessment videos

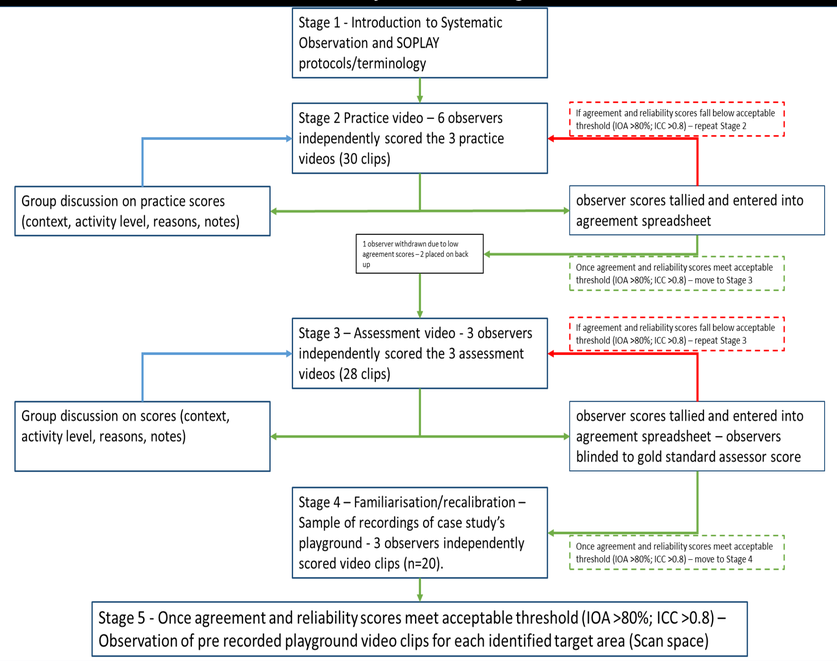


Figure 5. Stages of training observers and recording reliability data

Challenges and solutions

Systematic Observation

There are now various, validated and reliable methods for measuring PA (accelerometry, pedometry, self report, direct observation, heart rate monitoring, direct/indirect calorimetry).[6] However, many methods fail to consider the importance of the context in which PA occurs (McKenzie and van der Mars 2015). Direct, systematic observation exceeds other methods in simultaneously capturing the social and physical context of PA.[2, 7] However, the research team encountered some methodological challenges with using SO in a primary school playground environment.

Live observations – During the first familiarisation session at the school, it became very apparent to the team that live observations in a primary school playground of children's activities levels was unrealistic due to the environmental complexities. Continuing with this approach would have resulted in an inaccurate representation of the participants behaviours.

Video recordings – This method resulted in a larger number of observations been possible for each scan area with less of a visible presence, and is considered an advantage in observation studies for the ability to resolve uncertainties in activity codes by reviewing the recordings over and over.[7] Furthermore video recordings have previously been demonstrated to have high degrees of agreement with live assessments.[8] The volume of video recordings allowed for the construction of a range of sample video clips, which were used in conjunction with previously available gold standard assessments (Figure 5**.** stage 3) to check for any observer error ("observer drift", "observer reactivity", "observer bias") and the need for recalibration (retraining) of observers, or the use of back up observers.

Participant reactivity – With any SO method there is potential for participants to behave differently in the presence of the observers/research team. In order to reduce reactivity the research team followed the following **instructions**:

- Informed the participants about general details of the project and the presence of the observers prior to any data collection.
- Made the participants aware that they were required to take part in their usual playground activities and ignore the presence of the research staff..
- Utilised the mapping of the playground as an initial familiarisation for participants to the research team.
- Used video cameras to reduce the need for as many observers and therefore a reduced presence of the researchers in the playground.
- Observers were asked to be as inconspicuous as possible and to inform the children *'we are busy at the minute but I can talk to you later'* to avoid reactive behaviour.[7]
- In order to further reduce the reactivity to the video camera, any recordings taken before school and during morning break were considered as further familiarisation and not used in the scoring of playground activity levels.[7]

In any instance where the camera affected the activity/behaviour of the participants, the decision was made to delete this observation from the recordings. A designated **'no observation zone'** was made available to the participants. This area and any activities within this area were excluded from observations.

1. Mckenzie TL., van der Mars H. (2015) Top 10 Research Questions Related to Assessing Physical Activity and Its Contexts Using Systematic Observation. *Res Q Exerc Sport*: 86; 13-29 DOI: 10.1080/02701367.2015.991264
2. McKenzie, T.L., (2016). Context matters: systematic observation of place-based physical activity. *Res Q Exer Sport*; 87(4): pp.334-341
3. Shrout, PE., Fleiss, JL. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*; 86(2): 420–428. doi:10.1037/0033-2909.86.2.420
4. McHugh ML. (2012) Interrater reliability: the kappa statistic. *Biomechanics Medica*; 22(3): 276-82
5. Malcata RM., Vandenbogaerde TJ., Hopkins WG. (2014). Using athletes' World rankings to assess countries' Performance. *Int J Sports Physiol Performance*; 9(1): 133–138. doi:10.1123/ijspp.2013-0014
6. Ndahimana D., Kim E.K. (2017) Measurement Methods for Physical Activity and Energy Expenditure: a Review. *Clinical Nutrition Research*: 6(2); 68–80 http://doi.org/10.7762/cnr.2017.6.2.68
7. McKenna J., Zwolinsky, S. (2015) *Evaluation of Project HE:RO* - Keighley. Centre for Active Lifestyles, Leeds Beckett University.
8. Darst PW., Zakrajsek DB., Mancini VH. (1989) *Analyzing physical education and sport instruction 2nd ed*. Human Kinetics Inc. Champaign, IL.
9. Boonzaaijer M, van Dam E, van Haastert IC, Nuysink J. (2017) Concurrent validity between live and home video observations using the Alberta Infant Motor Scale. *Pediatr Phys Ther*; 29(2):146